

R Workshop Special Topic: Split Plot Analysis in R

Sarah Janse (sarah.janse@uky.edu)

Department of Statistics, University of Kentucky

November 4, 2016

Motivating Example (Agricultural Experiment)

We motivate our discussion by considering an agricultural investigation to study the effects of two irrigation methods (factor A) and two fertilizers (factor B) on yield of a crop. Let A1 and A2 be the two levels of factor A and B1 and B2 be the two levels of factor B. Four fields are available as the EUs. The four treatments - A1B1, A1B2, A2B1, and A2B2 - are assigned at random to the four fields in a CRD. Since there are four treatments and four experimental units (EUs), there will be no df for estimation of error as shown in the abbreviated ANOVA table (Table 1). However, if the fields could be divided into smaller fields, replicates of each factor-level combination could be obtained and the error variance could then be estimated. In this case, the smaller fields become the measurement units (MUs). Unfortunately, in this investigation, it is not possible to apply different irrigation methods (factor A) in areas smaller than a field (the EU), although different fertilizer types (factor B) could be applied in relatively small areas. Such a situation is where a split-plot design can be used.

Source of Variation	df
Factor A	1
Factor B	1
AB Interaction	1
Error	0
Total	3

Table 1: ANOVA table for example experiment

As noted in the motivating example, one factor sometimes requires more experimental material for its evaluation than a second factor in factorial experiments. In agricultural studies, a factor - such as irrigation method - may only be applicable (or best-suited) for large plots, while another factor - such as fertilizers - may be applied easily to smaller plots.

The larger treatment is applied to what we call the **whole plots**, which is then split into smaller **subplots**. Designs that accommodate this allocation of treatments are called split-plot designs. The usage of the term plots stems from split-plot designs being developed for agricultural studies; while still commonly found in agriculture, split-plot designs are also used in laboratory, industrial, and social science experiments. The primary advantage of a split-plot design is that it allows us to design an experiment when one factor requires considerably more experimental material than another factor, or accommodate the situation where there is an opportunity to study responses to a second factor while efficiently utilizing resources. The primary disadvantage of a split-plot design is the potential loss in treatment comparisons and an increase in complexity of the statistical analysis (which we will see). In split-plot designs, the usual randomization of treatment combinations to the EUs is altered to accommodate the particular requirements of the experiment. In the usual CRD, all treatment combinations are assigned randomly to the EUs. However, in the split-plot design, each level of the second factor is combined with a single level of the first factor (i.e., the factor requiring more experimental material), which has been assigned to the whole plot. In other words, at the first level of randomization, the whole plot treatments are randomly assigned to whole plots; at the second level, the subplot treatments are randomly assigned to the subplots.

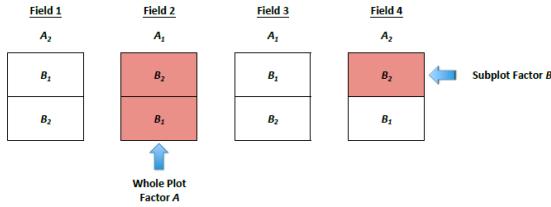


Figure 1: Whole plots and subplots layout

Example Experiment (Cont.)

As we noted in the agricultural experiment, the irrigation methods (factor A) can only be applied to one field as a whole. In Figure 1, we see that the four fields have been randomly assigned to one of the two factor levels of A. Within each field (the whole plot), it is divided into two subplots. Each of these subplots is then randomly assigned to one of the two levels of factor B. Through this split-plot design, we have increased our total number of EUs N from 4 to 8. In the analysis of split-plot designs, we must account for the presence of two different sizes of EUs used to test the effect of the whole plot treatment and subplot treatment. Factor A effects are estimated using the whole plots and factor B and the AxB interaction effects are estimated using the subplots. Since the size of whole plot and subplots are different, they have different precisions. Moreover, two separate errors account for the fact that observations from different subplots within each whole plot may be positively correlated, since it is natural to assume that EUs adjacent to one another are likely to respond similarly. For example, consider neighboring field plots, students in a classroom, or the batch of some raw material in an industrial experiment.

Thus, for the ANOVA, we will partition our experimental error into two components: one for the whole-plot treatment factor and one for the subplot treatment factor and interaction.

First Split Plot Example with CRD for Whole Plot

Consider an experiment involving the water resistant property of wood. Two types of wood pretreatment (one and two) and four types of stain (one, two, three and four) have been selected as variables of interest. Conducting this experiment in a completely randomized fashion would require eight wood panels for each full replicate of the design. Each wood panel would be randomly assigned a particular pretreatment and stain combination.

But it turns out to be very difficult to apply the pretreatment to a small wood panel. The easiest way would be to apply each of the pretreatment types (one and two) to an entire board, then cut each board into four smaller pieces and apply the four stain types to the smaller pieces. This is shown in Figure 2.

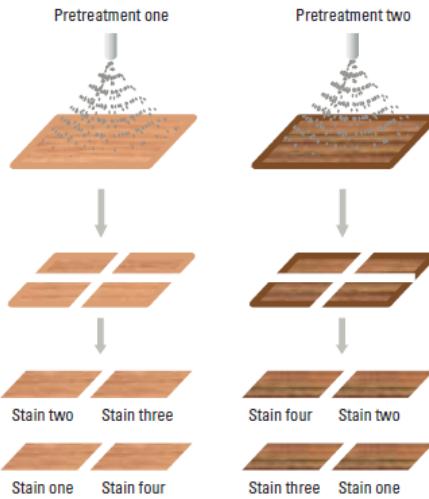


Figure 2: Treatment Application Layout

The reader should recognize this is a split-plot design for four reasons:

1. For the pretreatment factor, an experimental unit is the entire board or a set of four pieces of the board after they are cut. For the stain factor, an experimental unit is an individual piece of the board. Having unequal sized experimental units for the different factors is one key element of a split-plot design.
2. Each factor uses a different randomization scheme. In contrast, a complete randomized design would use one randomization scheme for all 24 experimental runs.
3. Note for a single run at one level of pretreatment, four separate runs are conducted for the stains. As a result, pretreatment could be thought of as a hard-to-change factor, while stain could be considered an easy to change factor.
4. The number of experimental replicates is not the same for each factor. Pretreatment has only three experimental replicates for each of the two factor levels, while stain has six experimental replicates for the stain factor levels.

```
#####
## Read in data
#####

## One option: Use the read.table() command
woodata = read.table('woodata.csv', # Give filename first
                      header=TRUE, # If filename has variable names, set header to TRUE.
                      # Otherwise, use header=FALSE
                      sep=",", # Symbol separating data values (comma here)
                      na.strings="NA")
## You can also read in the data using the Import Dataset option as we have previously.

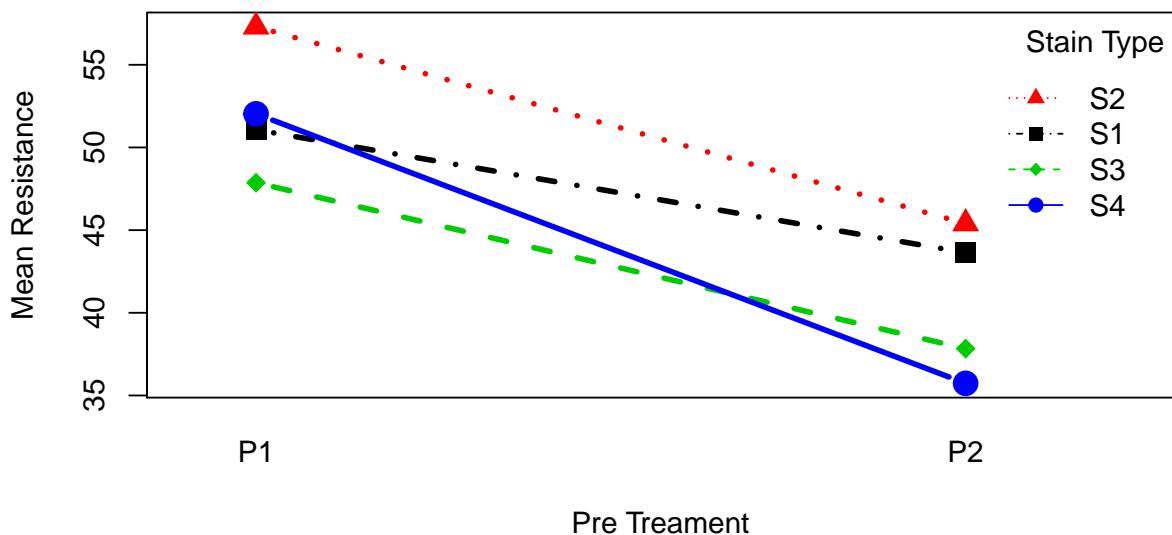
## First, we want to edit the data to make Rep a factor
woodata$Rep <- as.factor(woodata$Rep)
```

```

interaction.plot(x.factor=wooddata$Pre,
                  trace.factor=wooddata$Stain,
                  response=wooddata$Resistance,
                  trace.label='Stain Type', xlab='Pre Treatment',
                  ylab='Mean Resistance',
                  col=1:4,lwd=3,type='b',pch=c(15,17,18,19),
                  main='Factor Plot for Wood Example',cex=1.5
)

```

Factor Plot for Wood Example



```

out <- aov(Resistance~Pre*Stain+Error(Rep) ,data=wooddata)
summary(out)

##
## Error: Rep
##           Df Sum Sq Mean Sq F value Pr(>F)
## Pre         1  782.0   782.0   4.034  0.115
## Residuals  4  775.4   193.8
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Stain       3 266.01   88.67   6.976 0.00569 **
## Pre:Stain   3  62.79   20.93   1.647 0.23091
## Residuals 12 152.52   12.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Second Split Plot Example With RCBD for Whole Plot

A soil scientist wants to investigate the effects of nitrogen supplied in different chemical forms and evaluate those effects combined with those of thatch accumulation on the quality of turfgrass. Four forms of nitrogen fertilizer are used: urea, ammonium sulphate, isobutylidene diurea (IBDU), and sulphur-coated urea, urea (SC). A golf green was constructed and seeded with Penncross creeping bent grass on the experimental plots. The nitrogen treatment plots were arranged on the golf green in an RCBD with two replications. After two years, the second treatment factor - years of thatch accumulation - was added to the experiment. Each of the eight experimental plots was split into three subplots to which levels of the second treatment factor were randomly assigned. The lengths of time the thatch was allowed to accumulate on the subplot were 2, 5, or 8 years. Thus, the split plot design had the whole-plot treatment factor of nitrogen source in an RCBD with years of thatch accumulation as the subplot treatment factor. The measurements made on the turfgrass plots is the chlorophyll content of clippings (mg/g).

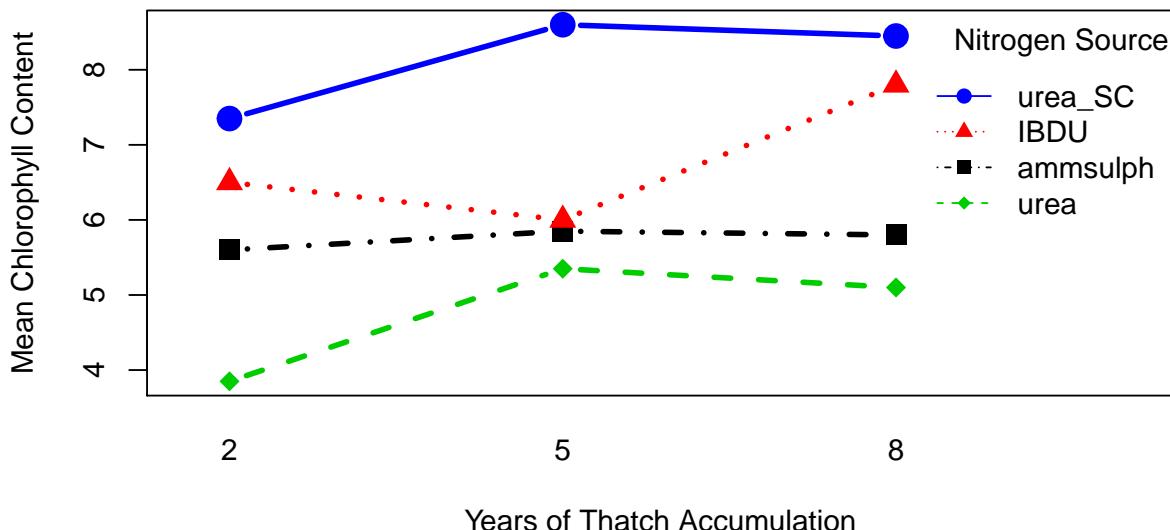
```

nit <- as.factor(c(rep("urea",6),rep("ammsulph",6),rep("IBDU",6),
                   rep("urea_SC",6)))
blk <- as.factor(c(rep(c(rep(1,3),rep(2,3)),4)))
years <- as.factor(rep(c(2,5,8),8))
content <- c(3.8,5.3,5.9,3.9,5.4,4.3,5.2,5.6,5.4,6.0,6.1,6.2,
            6.0,5.6,7.8,7.0,6.4,7.8,6.8,8.6,8.5,7.9,8.6,8.4)
field <- as.factor(sort(rep(1:8,3)))

grass <- data.frame(Nitrogen=nit,Block=blk,Years=years,
                     Content=content,Field=field)
interaction.plot(x.factor=grass$Years,trace.factor=grass$Nitrogen,
                  response=grass$Content,trace.label="Nitrogen Source",
                  xlab="Years of Thatch Accumulation",ylab="Mean Chlorophyll Content",
                  col=1:4,lwd=3,type="b",pch=c(15,17,18:19),
                  main="Factor Plot for Grass Experiment",cex=1.5)

```

Factor Plot for Grass Experiment



```

out <- aov(Content~Block+Nitrogen*Years+Error(Field), data=grass)
summary(out)

##
## Error: Field
##           Df Sum Sq Mean Sq F value Pr(>F)
## Block      1   0.51   0.510   1.217 0.3505
## Nitrogen   3  37.32  12.442  29.672 0.0099 **
## Residuals  3   1.26   0.419
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Years       2   3.816   1.9079   8.891 0.00927 **
## Nitrogen:Years 6   4.154   0.6924   3.227 0.06460 .
## Residuals    8   1.717   0.2146
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```